

Provendo recursos de processamento por meio de nuvem IaaS para subsidiar pesquisas agropecuárias

Jorge Luiz Corrêa¹ e Marcos Cezar Visoli¹

¹Embrapa Informática Agropecuária
Caixa Postal 6041 – 13083-886 – Campinas – SP – Brazil

{jorge,visoli}@cnptia.embrapa.br

Abstract. *This paper describes the implementation of an IaaS (Infrastructure as a Service) private cloud computing scenario at an Agricultural Research Corporation, using OpenStack. The paper discusses the main characteristics of cloud computing systems and OpenStack software. Moreover, shows two case studies of Juju's for service orchestration: building a Hadoop cluster and the establishment of a service (Wordpress blog). In general, discusses how these technologies have been used to support agricultural research.*

Resumo. *Este artigo descreve a implementação de um cenário de nuvem computacional privada IaaS (Infrastructure as a Service), dentro de uma instituição de pesquisa agropecuária, utilizando o OpenStack. O trabalho discute as principais características de sistemas de nuvem computacional e da pilha de software OpenStack. Ainda, mostra dois estudos de caso da utilização do Juju para orquestração de serviços: construção de um cluster Hadoop e estabelecimento de um serviço web (blog Wordpress). De modo geral, discute como estas tecnologias tem sido utilizadas no auxílio a pesquisas agropecuárias.*

1. Introdução

A computação em nuvem é uma tecnologia recente dentro da computação que tem ganhado grande relevância nos últimos anos. Diversos fatores tem impulsionado este ganho de importância, como as características técnicas que esta tecnologia pode fornecer, como por exemplo, elasticidade, otimização de recursos de hardware, organização de *datacenters*, orquestração de serviços, ou mesmo em outros contextos, como na área administrativa das instituições, com características como a taxaço pelo uso, conversão de custos de capitais em custos operacionais, menores investimentos em infraestrutura e não depreciação dos valores de ativos de hardware. Embora o termo 'computação em nuvem' não seja novo, apenas recentemente os usuários puderam conhecer mais detalhadamente a tecnologia. Isto se deve principalmente ao aumento de empresas fornecendo serviços em nuvem e ao significativo avanço das plataformas de software utilizadas para a construção de nuvens.

Segundo o NIST (*National Institute of Standards and Technology*) computação em nuvem “é um modelo para permitir acesso via rede, conveniente e sob demanda, a um conjunto de recursos computacionais configuráveis e compartilhados (como redes,

servidores, armazenamento, aplicações e serviços) que podem ser rapidamente provisionados e liberados com o mínimo de esforço de gerenciamento e interação com o provedor do serviço” [NIST 2012]. Este modelo é majoritariamente utilizado por empresas que vendem serviços, como processamento (servidores) e armazenamento, pois fornece grande agilidade em TI combinada a uma grande redução de custos. Com o surgimento de plataformas de software que permitem a implementação de uma nuvem própria, este modelo passou a ser estudado como forma de provimento de serviços internos nas instituições.

Diversas são as vantagens que uma estrutura de computação em nuvem pode fornecer. Dentre as principais estão:

- Agilidade no provisionamento de hardware: geralmente, quando há a demanda de um novo servidor, é necessária a alocação de espaço físico e transporte do hardware; com o estabelecimento de uma nuvem, os servidores criados são na maioria virtuais, executando em *hosts* do *datacenter*, e acessados remotamente;
- Agilidade no provisionamento de sistemas: as imagens de sistemas estão prontas para serem usadas, não sendo necessárias repetidas instalações e configurações de sistemas operacionais e dos software mais utilizados;
- Elasticidade: se determinado experimento demanda mais capacidade de processamento ou armazenamento, esta pode ser concedida ao usuário criando-se novas instâncias de servidores ou volumes de armazenamento;
- Melhor aproveitamento dos recursos computacionais: uma vez que um servidor virtual está ocioso ele pode ser liberado, disponibilizando o hardware para outros projetos, rapidamente;
- Economia de energia: a virtualização permite que o poder computacional de cada *host* seja melhor utilizado de modo que *hosts* não fiquem ligados e ociosos; além disso, uma quantidade menor de *hosts* representa menor gasto com refrigeração e, conseqüentemente, menor gasto com energia.

Este trabalho apresenta uma proposta de como se utilizar os benefícios e inovações da computação em nuvem aplicados em uma instituição de pesquisa agropecuária do governo brasileiro.

2. Contextualização

Este trabalho se desenvolve no contexto de uma empresa governamental brasileira de pesquisa agropecuária, a Embrapa (Empresa Brasileira de Pesquisa Agropecuária). A Embrapa é uma empresa cuja missão é viabilizar soluções de pesquisa, desenvolvimento e inovação para a sustentabilidade da agricultura, em benefício da sociedade brasileira. Atua por intermédio de unidades de pesquisas em diversas áreas da agropecuária, presentes em todo o território nacional. O trabalho apresentado neste artigo foi desenvolvido em uma unidade específica de tecnologia da informação para a agropecuária (Embrapa Informática Agropecuária) que atua em diversos projetos nos quais a computação é utilizada de forma essencial.

Existem diversas linhas de pesquisa cujos projetos utilizam recursos

computacionais, como sistemas voltados para a pesquisa com bancos de dados, mineração de dados, geotecnologia, modelagem agroambiental, estudos climáticos, inteligência computacional e bioinformática. Muitas dessas atividades demandam constantemente recursos de processamento e armazenamento de informações.

O modelo atualmente utilizado para provisionamento desta infraestrutura é manual, de forma que ao surgir a demanda por parte de pesquisadores, deve-se efetuar uma solicitação que terá sua possibilidade analisada. Em caso de haver a possibilidade, o hardware deve ser alocado em algum espaço físico. Após a instalação física, inicia-se a instalação lógica, ou seja, sistemas operacionais e software que serão utilizados. Todo este processo demora, normalmente, alguns dias, principalmente se não houver disponibilidade de recursos computacionais. E, muitas vezes, estes recursos não estão disponíveis por já estarem alocados a outros projetos, porém, ociosos.

Um dos objetivos deste trabalho é justamente utilizar as inovações na maneira de se prover serviços oferecidas pelas nuvens para auxiliar as pesquisas da unidade. A agilidade no gerenciamento de infraestrutura de TI pode impactar na produtividade científica, uma vez que torna muito mais versátil todas as questões de gerenciamento de hardware e software, de modo que o tempo seja aplicado de fato nas atividades de pesquisa, desenvolvimento e inovação tecnológica.

3. Trabalhos relacionados

As bases conceituais que possibilitaram o estabelecimento das nuvens computacionais não são novas, estando relacionadas com temas já difundidos tais como sistemas distribuídos, computação em grade, virtualização e *outsourcing*. A computação em nuvem nos moldes atuais é uma convergência desses conceitos e tecnologias [Taurion 2009]. Porém, o uso de uma nuvem computacional com a visão de serviços é relativamente novo. As ferramentas gratuitas (código aberto) que possibilitam esta utilização estão, em grande parte, ainda em desenvolvimento e aprimoramento (em processo de adoção pela comunidade). Assim, o uso de nuvens computacionais tem sido predominantemente comercial, auxiliado por ferramentas proprietárias. Grandes empresas fazem uso desta nova tecnologia para a venda de serviços. Um novo modelo de negócios surgiu pautado na venda de serviços por demanda, com formas diferenciadas de tarifação. O exemplo mais conhecido deste modelo é o criado pela Amazon, em seus serviços EC2 [Amazon EC2 2012] e S3 [Amazon S3 2012], onde usuários podem comprar processamento e armazenamento, respectivamente. Além disso, também são oferecidos como serviços ambientes completos de desenvolvimento, baseados em WEB, com hospedagem na própria empresa fornecedora do serviço. Os exemplos mais conhecidos são o [Google App Engine 2012] e o [Windows Azure 2012].

Como auxílio à pesquisa científica, destaca-se a utilização do Nimbus [Nimbus 2012]. Trata-se de um projeto de software livre utilizado por entidades científicas. Assim como outras ferramentas abertas, está em constante desenvolvimento, como mostra o trabalho sobre a eficiência do armazenamento em nuvem oferecido pelo Nimbus [Bresnahan, J. et. al. 2011].

Existem também trabalhos mostrando a utilização de outros sistemas em

conjunto com estruturas chamadas de nuvem por seus autores. Dois exemplos são [Hyeokju Lee et. al. 2012] e [Hao, Chen et. al. 2011]. O primeiro mostra como foi implementada uma estrutura para conversão de imagens utilizando o Hadoop e uma biblioteca Java. O segundo mostra um estudo sobre o Hadoop, mais especificamente sobre o sistema de armazenamento distribuído, denominado HDFS. Ambos os autores, assim como em diversos outros trabalhos, mencionam estruturas de nuvens computacionais. No entanto, a exemplo do trabalho de Hyeokju Lee, foi utilizado um *cluster* computacional e não uma estrutura real de nuvem.

Uma vez que os primeiros *frameworks* de computação em nuvem começaram a surgir de maneira mais estável apenas recentemente, são poucos os trabalhos neles baseados. Neste aspecto, é possível utilizar a computação em nuvem, a partir de um *framework* próprio para construção de nuvens computacionais, aplicada no auxílio ao desenvolvimento de pesquisas agropecuárias.

4. Nuvens computacionais

As nuvens computacionais podem ser categorizadas de duas maneiras. Primeiro, referente ao tipo de uso e acesso. Neste contexto elas podem ser públicas ou privadas. Nuvens públicas normalmente são estruturas de empresas que vendem serviços e são acessadas por meio da Internet. Todas as atividades são realizadas em zonas mantidas por estas empresas, podendo ser replicadas em diversas localizações. Já as nuvens privadas são utilizadas por instituições a fim de prover os mesmos serviços comercializados em nuvens públicas, porém internamente, com acesso restrito.

Outra maneira de categorizar as nuvens é quanto ao tipo de serviço fornecido: SaaS (*Software as a Service*), PaaS (*Platform as a Service*) ou IaaS (*Infrastructure as a Service*). Nuvens SaaS disponibilizam um software como serviço. Assim, o usuário acessa uma aplicação, como por exemplo um editor de textos e planilhas, normalmente via web, que estará sendo executada em algum lugar dentro da nuvem. Uma nuvem PaaS fornece ao usuário uma plataforma de desenvolvimento. Usuários podem então desenvolver novas aplicações em um ambiente que é executado remotamente e acessado por meio de um ponto de acesso da nuvem. Por fim, nuvens IaaS fornecem infraestrutura como serviço, como por exemplo, processamento e armazenamento. Os usuários de nuvens IaaS criam computadores virtuais, em quantidades e capacidades demandadas e os acessam remotamente. Outro serviço IaaS é o armazenamento, de modo que o usuário utiliza a nuvem como forma de armazenar dados, sendo que as principais características associadas a esta atividade na nuvem são a redundância e disponibilidade dos dados. Neste sentido, pretende-se suprir a demanda por dois serviços essenciais para a realização das pesquisas: provisionamento ágil de infraestrutura de processamento e de armazenamento confiável.

5. Nuvem privada IaaS com OpenStack

O OpenStack é um conjunto de ferramentas livres que possibilita a criação de uma nuvem computacional IaaS, incluindo serviços de armazenamento de dados. Os maiores contribuidores para que o OpenStack atingisse o grau de desenvolvimento atual foram a empresa Rackspace Hosting e a NASA. Atualmente o projeto tornou-se uma comunidade de software global, com diversos colaboradores, dentre eles grandes

empresas como a Dell, HP, Citrix, Intel, AMD e a Canonical.

O OpenStack é organizado em três famílias de serviços: *Compute Infrastructure* (Nova), *Storage Infrastructure* (Swift) e *Imaging Service* (Glance). Cada uma dessas famílias possui um ou mais utilitários responsáveis por desempenhar determinada função dentro da nuvem.

5.1. Compute infrastructure (Nova)

Este conjunto de ferramentas é o controlador da nuvem. Todas as atividades necessárias para que um instanciamento de máquina virtual ocorra são gerenciadas por esta entidade como, por exemplo, recursos computacionais, rede, autorização e escalabilidade. No entanto, trata-se de uma plataforma de gerenciamento, não existindo nenhuma capacidade de virtualização. Todo o processo de virtualização no OpenStack é possibilitado por algum dos *hypervisors* suportados, de modo que a gerência do processo de virtualização é executado na plataforma Nova. Esta gerência é baseada na API da biblioteca *libvirt*, utilizada por diversos sistemas de virtualização. Segundo a documentação atual do OpenStack, são suportados o Xen, XenServer, KVM, UML e Hyper-V. Dentro do conjunto OpenStack Infrastructure Compute, existem componentes específicos para a realização de diversas tarefas. São eles:

API Server: responsável por traduzir chamadas do mundo externo para ações dentro da nuvem por meio de chamadas Web Services.

Message queue: sistema que implementa o protocolo AMQP (*Advanced Message Queue Protocol*) e controla a troca de mensagens entre os componentes Nova.

Compute worker: é o componente que dá origem a um nó de computação, sendo gerenciador do ciclo de vida das instâncias virtuais.

Network controller: componente responsável por implementar os controles de rede dentro da infraestrutura de nuvem, gerenciando tanto operações nos nós da nuvem quanto nas máquinas virtuais.

Volume worker: gerencia os discos virtuais para armazenamento permanente. No OpenStack, uma instância possui armazenamento volátil. Isto significa que, quando ela for desalocada da nuvem, os dados de seu sistema de arquivos serão perdidos. Para prover armazenamento persistente são utilizados volumes lógicos que são montados nas instâncias virtuais. Este componente é quem gerencia estes volumes, permitindo que instâncias armazenem dados em um sistema de arquivos permanente.

Scheduler: este componente escalona ações dentro da nuvem. Uma ação importante é a definição de qual nó físico da nuvem executará uma instância virtual. Ele possui algoritmos que permitem regular como a escolha ocorrerá, considerando fatores como carga de processamento, memória disponível e arquitetura da CPU.

5.2. Storage infrastructure (Swift)

O Swift é uma infraestrutura de armazenamento de dados na forma de objetos. Deve ficar claro que o Swift é diferente do *Volume Worker*, um componente do OpenStack responsável por possibilitar que discos virtuais sejam utilizados nas instâncias. O Swift

pode ser entendido como um sistema capaz de fornecer armazenamento como serviço. Uma vez que seu objetivo é fornecer armazenamento, diversas características são inerentes de sua utilização, como por exemplo, gerenciamento de redundância e *failover*, capacidade de servir gráficos e vídeos (*stream*), escalabilidade, dentre outras. Trata-se, portanto, de uma maneira de se armazenar dados na nuvem com diversas características de suporte e otimização relativas a tal atividade, buscando abstrair para o usuário toda a complexidade do processo de custódia da informação armazenada.

5.3. Imaging Service (Glance)

O Glance é um serviço de pesquisa e recuperação de imagens de máquinas virtuais. Além da tarefa de gerenciamento, dentro do OpenStack ele pode ser entendido como o repositório das imagens das máquinas virtuais. A utilização da virtualização possibilita que diversos sistemas operacionais possam ser utilizados. Na estrutura do OpenStack, é possível gerar imagens virtuais contendo cada um destes sistemas. Estas imagens são armazenadas no Glance e, quando um usuário da nuvem requisita uma instância, ele deve informar qual imagem será utilizada para gerar a instância virtual. Portanto, o Glance permite o gerenciamento destas imagens, definindo formatos, metadados e realizando o armazenamento, de fato.

Visando suprir as demandas anteriormente mencionadas, a arquitetura de nuvem mostrada na figura 1 foi elaborada, utilizando os componentes do OpenStack.

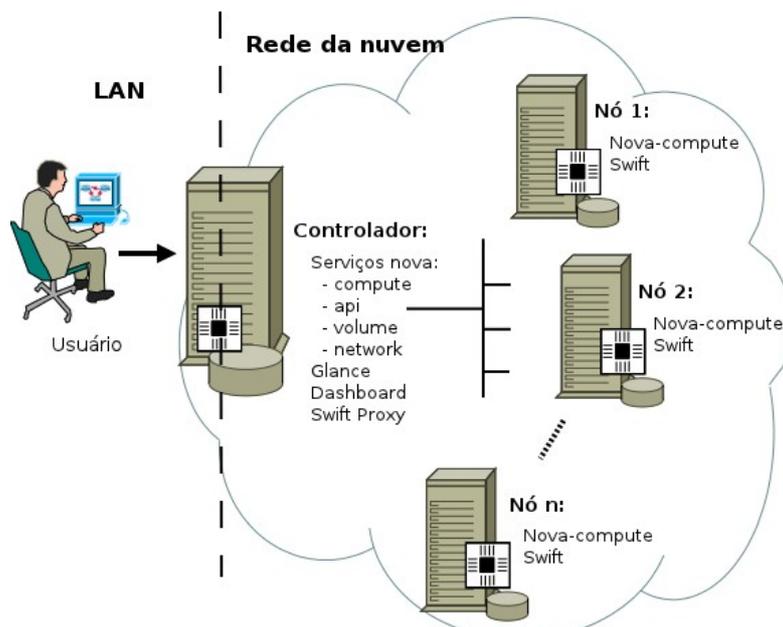


Figura1. Arquitetura da nuvem computacional privada baseada no OpenStack.

6. Orquestração de serviços com o Juju

O Juju é uma ferramenta livre que atua em conjunto com uma nuvem baseada no OpenStack, permitindo a orquestração de serviços. A orquestração possibilita que as melhores práticas de configuração de serviços sejam compartilhadas e reusadas. No contexto do Juju, estas melhores práticas são traduzidas em *scripts* (denominados *charms*) que guiam o estabelecimento de um serviço. Em conjunto com uma nuvem

privada estabelecida com o OpenStack, o Juju permite o estabelecimento de diversos serviços, desde o sistema operacional, pela criação de instâncias virtuais, até a configuração lógica das aplicações, como bancos de dados, sistemas de gerenciamento de conteúdo, e diversos outros serviços. O Juju fornece uma solução simples para que um conjunto de serviços seja criado, configurado e integrado operacionalmente.

7. Aplicação na pesquisa agropecuária

Conforme mencionado, o desenvolvimento deste trabalho se dá no contexto de uma empresa de pesquisa agropecuária. Existem diversas aplicações da computação neste tipo de pesquisa, principalmente para a realização de simulações com modelos matemáticos, modelagens agroambientais, mineração de textos, processamento de informações genéticas, imagens e de dados geoespaciais. Além disso, muitas vezes são gerados sistemas para usuários finais, tais como sistemas web que utilizam mecanismos e modelos computacionais criados a partir das pesquisas. Neste contexto, serão mostrados dois exemplos de uso de uma nuvem IaaS implementada com o OpenStack, que se relacionam diretamente com estas atividades.

7.1. Criação de um *cluster* Hadoop para processamento distribuído

O Hadoop (Hadoop, 2012) é um *framework* para processamento distribuído de informações. Ele permite que grandes quantidades de dados sejam armazenadas em um sistema de arquivos distribuído e que, colaborativamente, diversos nós realizem processamentos sobre esta massa de dados. O Hadoop trabalha com as operações de Map, em que há um mapeamento do conjunto inicial de dados em porções menores, e Reduce, em que os resultados dos processamentos das porções são agrupados para produzir uma saída única. O processamento em si é definido pelo usuário. O Hadoop permite que a lógica das operações de Map e Reduce sejam criadas pelo pesquisador, de modo a atender suas necessidades para a resolução de um problema. Dada a versatilidade no processamento de informações que o Hadoop oferece, a criação de *clusters* Hadoop torna-se bastante relevante.

Inicialmente foi realizada a instalação do software Juju (Juju, 2012) para a orquestração de serviços. Uma vez que o OpenStack utiliza o sistema Keystone para autenticação dos usuários, novas credenciais foram criadas para que o Juju pudesse interagir com a nuvem. O arquivo de configuração resultante é mostrado a seguir (~/.juju/environments.yaml).

```
environments:
  sample:
    type: ec2
    default-instance-type: m1.small
    control-bucket: juju-bucket
    admin-secret: admin
    access-key: d8c193a62ff24d5f947b3bbf8a86d9c9
    secret-key: 1d99c7bcbbfa420faf92347427f77251
    ec2-uri: http://10.129.10.44:8773/services/Cloud
    s3-uri: http://10.129.10.44:3333
    default-image-id: oneiric-11.10-hadoop-image
    ec2-key-name: root44
    juju-origin: distro
```

```
default-series: oneiric
```

Este arquivo possui as credenciais para que o Juju interaja com a nuvem e parametrize o ambiente "sample" para utilizar o tipo de instância "small" (2 GB de RAM e 1 Vcpu) cuja imagem do sistema de arquivos é "oneiric-11.10-hadoop-image". Esta imagem possui configurações de *proxy* e do sistema de pacotes APT (*mirror* local). Por fim, uma chave SSH denominada *root44* é inserida nas instâncias por padrão.

O Juju trabalha com os chamados *charms*, que são conjuntos de instruções para a instalação, inicialização, configuração, estabelecimento de relacionamentos e paralisação de um serviço. Os *charms* são *scripts* que seguem uma estrutura padrão. Para a criação de um *cluster* Hadoop foram utilizados dois *charms* obtidos de (Hadoop Charms, 2012): *hadoop-master* e *hadoop-slave*. Algumas seções destes *scripts* foram modificadas para solucionar problemas relativos ao ambiente local. O *hadoop-master* cria uma instância virtual que controlará o restante dos nós, também conhecida como *namenode*. Já o *charm* *hadoop-slave* cria uma instância chamada *datanode*, que servirá para armazenamento de dados e processamento. A seguir são mostrados as chamadas responsáveis por iniciar a criação do *cluster*.

```
juju bootstrap
juju deploy --repository ~/charms local:hadoop-master
juju deploy --repository ~/charms local:hadoop-slave
juju add-relation hadoop-slave hadoop-master
juju expose hadoop-master
```

O comando de *bootstrap* faz com que uma instância de controle seja criada dentro da nuvem. É a partir desta instância que o Juju orquestrará a criação de novas máquinas virtuais e configurações de seus serviços. Os dois comandos a seguir executam a implantação de uma instância *master* e outra *slave*, utilizando um repositório local de *charms*, em *~/charms*. Na estrutura utilizada, o instanciamento destas duas máquinas virtuais, bem como a instalação dos software necessários, configuração e inicialização, demora cerca de 3 minutos. O nó controlador possui um processador de 4 núcleos com 3 GHz, 8 GB de memória RAM, disco SATA 5200 RPM e rede Gigabit. O nó de processamento possui 2 processadores de 6 núcleos com 3 GHz, 64 GB de memória RAM, discos SAS 7200 RPM e rede Gigabit. Uma vez funcionais, o comando *add-relation* cria uma relação entre os dois nós. Em uma configuração manual do Hadoop, seria necessário alterar o *master* para informar a existência de um novo *slave*, indicando principalmente seu endereço IP, além de outras configurações, e o próprio *slave*, semelhantemente. Utilizando orquestração com o Juju, todas estas configurações são automáticas. Por fim, o comando *expose* libera acesso ao nó *master* instanciado dentro da nuvem. A infraestrutura do OpenStack utiliza uma rede privada para as instâncias virtuais. Todo acesso às instâncias é restrito e controlado pelo Network Controller. O comando *expose* faz com que o acesso à instância seja liberado, sendo que a liberação é realizada de acordo com o serviço que o Juju criou na máquina virtual. Assim, neste caso, serão liberadas apenas as portas utilizadas pelo Hadoop, como por exemplo a porta 50070, onde se pode acessar uma interface de visualização do *cluster*. Quanto ao endereçamento, a nuvem tem à disposição um *pool* de endereços da LAN e foi configurada para atribuir um destes endereços para cada nova instância.

Os passos realizados até então criam, de fato, uma instância controladora e um

nó de processamento. Uma vez que o objetivo é criar um *cluster* Hadoop, com vários nós de processamento, o Juju permite que estes novos nós sejam adicionados facilmente. Para fins de demonstração, foram criados apenas mais três nós do tipo *slave*. Esta adição de nós é realizada pelo comando *add-unit*. Deve ficar claro que os comandos de adição de relação e novas unidades devem possuir suas ações nos *scripts* do *charm*, como é o caso dos *charms* do Hadoop utilizados. Para a criação de novos nós, pode-se utilizar a chamada a seguir, alterando o laço para a quantidade necessária.

```
for i in {1..3} ; do
    juju add-unit hadoop-slave;
done
```

Terminados os instanciamentos, o *cluster* estará operacional. O acesso às máquinas poderá ser realizado por meio de SSH utilizando a chave indicada na configuração do Juju. O endereço será algum IP do *pool* mencionado, de forma que o controlador de rede da nuvem automaticamente configura as operações de NAT (Network Address Translation) para os endereços internos. As figuras 2 e 3 mostram, respectivamente, as instâncias virtuais na interface de gerenciamento da nuvem e os nós do *cluster* Hadoop, sendo esta última uma interface na instância master (namenode).

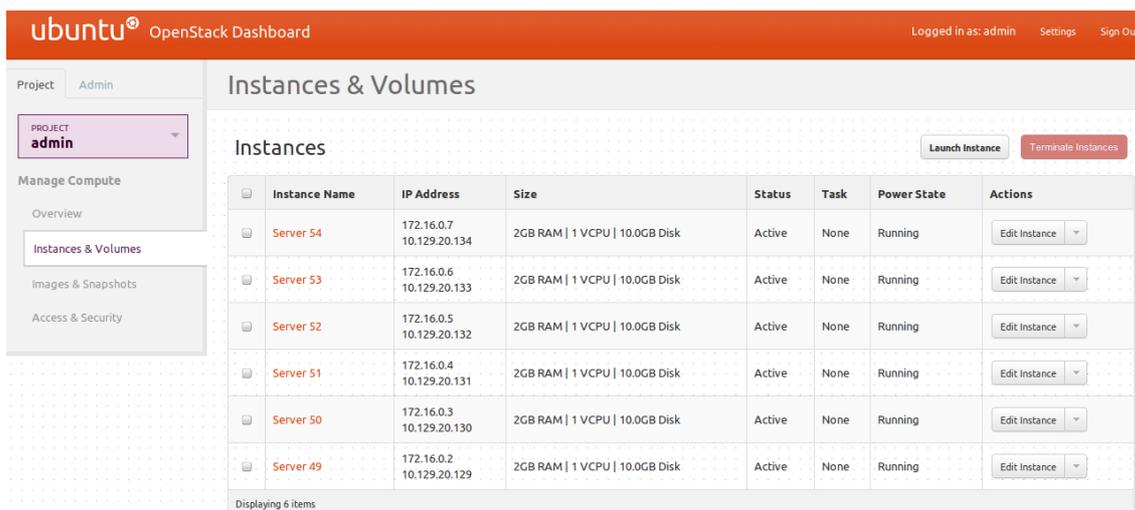


Figura 2. Interface de gerenciamento da nuvem, Dashboard, com as instâncias do cluster Hadoop.

NameNode 'server-50.novalocal:8020'

Started:	Fri May 04 20:35:20 UTC 2012
Version:	0.20.2-cdh3u1, rbdafb1dbffd0d5f2fbc6ee022e1c8df6500fd638
Compiled:	Mon Sep 19 21:04:59 UTC 2011 by buildd
Upgrades:	There are no upgrades in progress.

[Browse the filesystem](#)
[NameNode Logs](#)
[Go back to DFS home](#)

Live Datanodes : 4

Node	Last Contact	Admin State	Configured Capacity (GB)	Used (GB)	Non DFS Used (GB)	Remaining (GB)	Used (%)	Used (%)	Remaining (%)	Blocks	Failed Volumes
server-51	2	In Service	9.84	0	2.12	7.72	0	<div style="width: 0%;"></div>	78.42	1	0
server-52	2	In Service	9.84	0	2.12	7.72	0	<div style="width: 0%;"></div>	78.42	1	0
server-53	1	In Service	9.84	0	2.12	7.72	0	<div style="width: 0%;"></div>	78.42	1	0
server-54	1	In Service	9.84	0	2.12	7.72	0	<div style="width: 0%;"></div>	78.42	0	0

Figura3. Interface de visualização do cluster Hadoop na instância master.

Uma vez que é possível estabelecer diversos ambientes utilizando o Juju, ele possui uma ferramenta para geração de uma visualização gráfica dos mesmos. Como apenas este ambiente de *cluster* está sendo orquestrado, a figura gerada mostra as instâncias criadas e suas relações. Esta representação pode ser obtida com a seguinte chamada.

```
juju status --format svg --output hadoop.svg
```

A representação gerada pode ser vista na figura 4.

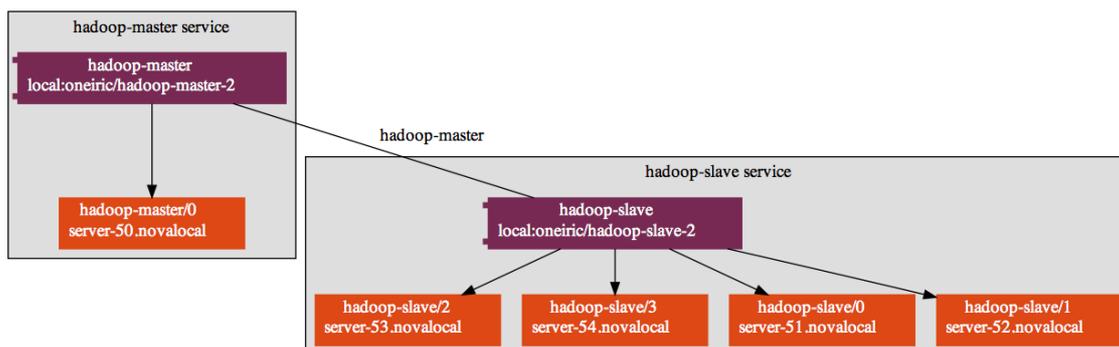


Figura4. Representação gráfica de um ambiente orquestrado pelo Juju.

7.2. Criação de um serviço web com banco de dados

Outra aplicação da orquestração de serviços é para a criação de sites web comuns, como gerenciadores de conteúdo, páginas wiki, forums e blogs. Um exemplo testado na infraestrutura de nuvem estabelecida foi a criação de um Wordpress, sistema de blog, utilizando um banco de dados MySQL. Os seguintes passos foram realizados para a obtenção deste ambiente.

```
juju bootstrap
juju deploy --repository ~/charms local:mysql
juju deploy --repository ~/charms local:wordpress
juju add-relation mysql wordpress
juju expose wordpress
```

Assim como no caso do Hadoop, todas as configurações são realizadas automaticamente. São criadas duas instâncias virtuais, uma para o MySQL e outra para o Wordpress. A seguir uma relação é criada para que o sistema de blog armazene suas informações no banco de dados em questão. Por fim, a instância com o Wordpress tem acesso liberado para usuários da LAN. A única diferença em relação ao Hadoop está na imagem utilizada para a criação das instâncias virtuais. Neste caso, foi utilizada uma imagem UEC contendo um Ubuntu Server 11.10 LTS, sem qualquer modificação. Imagens UEC são imagens próprias para execução em nuvens. Da mesma forma que a distribuição é liberada em formato .ISO, as imagens UEC (Ubuntu Cloud Guest) são distribuídas em formatos próprios para serem utilizadas em nuvens, como no OpenStack. Estas imagens podem ser obtidas em (UEC, 2012). Elas são equivalentes a uma instalação realizada com a imagem .ISO, com alguns pacotes próprios para a execução em nuvem, principalmente o pacote cloud-init.

8. Considerações finais e conclusão.

As nuvens computacionais tem se tornado uma realidade em ambientes de TI. Embora atualmente haja um grande tendência em se discutir nuvens públicas, fornecendo serviços para clientes com modelos de cobrança diferentes daqueles convencionais, as nuvens privadas também tem ganhado relevância.

É certo que ambientes empresariais privados serão muito beneficiados com o avanço desta tecnologia, pois uma de suas promessas é a redução de custos de capitais, transformando-os em custos operacionais. Pequenas e médias empresas que não investem em infraestruturas de TI poderão contratá-la de um CSP (*Cloud Service Provider*) a custos bem reduzidos, não necessitando manter funcionários para o suporte e não enfrentando a cíclica depreciação dos ativos de hardware.

No entanto, alguns cenários podem não se beneficiar destes serviços de baixo custo sendo disponibilizados no mercado. As instituições de pesquisas científicas são exemplos clássicos. Pesquisadores que trabalham com informações sigilosas não utilizarão infraestruturas externas às suas instituições para processamento e armazenamento de dados. No setor privado também haverá este tipo de restrição em instituições que mantêm segredos industriais ou informações com estas características. Neste contexto, as nuvens privadas podem fornecer grandes subsídios para a evolução da TI nestas instituições.

Neste trabalho, procurou-se mostrar como a utilização de uma nuvem para provimento de infraestrutura pode auxiliar a TI de uma instituição. A demanda por servidores é atualmente solucionada de maneira manual. Normalmente o processo envolve uma solicitação formal, a análise de possibilidade de se atender à solicitação, a alocação de espaço físico e a configuração lógica do sistema, pela instalação de sistemas operacionais e software. Com o uso de uma nuvem nos moldes da utilizada neste trabalho, o próprio pesquisador ou usuário que demandar infraestrutura poderá criá-la, diretamente, sem envolver etapas que atrasem o processo. A equipe de infraestrutura que atualmente realiza o processo manual ficará responsável pela manutenção da nuvem enquanto o próprio usuário irá interagir com ela, criando e liberando recursos.

Pela forma como o OpenStack é organizado, é possível realizar um controle bem apurado do uso dos recursos da nuvem. Pode-se criar usuários e projetos, de forma a atribuir cotas para cada um deles, como por exemplo, limitando a quantidade máxima de instâncias, de memória utilizada, de CPUs virtuais e de armazenamento. Desta forma, evita-se que determinado hardware, alocado para determinado projeto ou usuário, fique ocioso. Estando este hardware compondo uma estrutura de nuvem, quando de sua ociosidade, outros usuários poderão melhor utilizá-lo.

Embora não tenha sido mencionado detalhadamente, a infraestrutura criada também fornece o armazenamento como serviço. O componente do OpenStack responsável por prover armazenamento é o Swift. Conforme pode ser observado na figura 1, existe um Swift Proxy no nó controlador e cada um dos nós da nuvem fornece capacidades para compor um sistema único de armazenamento. O Swift permite que *containers* sejam criados e, dados neles armazenados, inerentemente contarão com replicação para manter a redundância e alta disponibilidade.

Uma vez que sistemas utilizados em atividades de pesquisa não apresentam as características que sistemas computacionais de uso comum necessitam, como estabilidade e alta disponibilidade, trata-se de um bom contexto para a obtenção de conhecimento e adaptação à nova tecnologia. Os atuais *frameworks* de computação em nuvem estão em pleno desenvolvimento e evolução. Espera-se como trabalhos futuros acompanhar este desenvolvimento a fim de aplicar a tecnologia na gestão de infraestrutura de modo geral, utilizando em conjunto com a virtualização para a melhor organização dos *datacenters* institucionais.

Referências

- AMAZON EC2. Elastic Compute Cloud. Pricing. Disponível em: <http://aws.amazon.com/ec2/#pricing>. Acesso em 2 de mai. 2012.
- AMAZON S3. Simple Storage Service. Pricing. Disponível em: <http://aws.amazon.com/s3/#pricing>. Acesso em 2 de mai. 2012.
- Bresnahan, J., LaBissoniere, D., Freeman, T., Keahey, K. ScienceCloud. Cumulus: An Open Source Storage Cloud for Science, 2011, San Jose, CA. June 2011.
- Google App Engine (Development Environment). Disponível em: <http://code.google.com/intl/pt-BR/appengine/docs/whatisgoogleappengine.html>. Acesso em 2 de mai. 2012.
- Hadoop. Apache Hadoop Project. Disponível em: <http://hadoop.apache.org/>. Acesso em 2 de mai. 2012.
- Hadoop Charms. Disponível em: <https://code.launchpad.net/charms>. Acesso em: 05 de mai. 2012.
- Hao, Chen; Ying, Qiao; , "Research of Cloud Computing Based on the Hadoop Platform," Computational and Information Sciences (ICCIS), 2011 International Conference on , vol., no., pp.181-184, 21-23 Oct. 2011.
- Hyeokju Lee; Myoungjin Kim; Joon Her; Hanku Lee; , "Implementation of MapReduce-based image conversion module in cloud computing environment," Information Networking (ICOIN), 2012 International Conference on , pp.234-238, 1-3 Feb. 2012.
- Juju. Disponível em: <https://juju.ubuntu.com/>. Acesso em: 05 de mai. 2012.
- NIMBUS. Open source toolkit for IaaS scientific community. Disponível em: <http://www.nimbusproject.org/>. Acesso em 2 de mai. 2012.
- NIST. Cloud Computing Program. Disponível em: <http://www.nist.gov/itl/cloud/>. Acesso em: 2 de mai. 2012.
- Taurion, Cezar. Cloud computing: computação em nuvem: transformando o mundo da tecnologia da informação. Rio de Janeiro: Brasport, p. 2, 2009.
- UEC. Ubuntu Cloud Guest. Disponível em: <https://help.ubuntu.com/community/UEC/Images>. Acesso em 05 de mai. 2012.
- Windows Azure. Ambiente Microsoft para desenvolvimento e hospedagem. Disponível em: <http://www.microsoft.com/windowsazure/pt/br/>. Acesso em 2 de mai. 2012.